



Extracting the Core Structural Connectivity Network: Guaranteeing Network Connectedness Through a Graph-Theoretical Approach

Demian Wassermann, Dorian Mazaure, Guillermo Gallardo-Diez, Rachid
Deriche

► To cite this version:

Demian Wassermann, Dorian Mazaure, Guillermo Gallardo-Diez, Rachid Deriche. Extracting the Core Structural Connectivity Network: Guaranteeing Network Connectedness Through a Graph-Theoretical Approach. MICCAI 2016, Sep 2016, Athens, Greece. hal-01333301v2

HAL Id: hal-01333301

<https://inria.hal.science/hal-01333301v2>

Submitted on 12 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extracting the Core Structural Connectivity Network: Guaranteeing Network Connectedness Through a Graph-Theoretical Approach

Demian Wassermann¹, Dorian Mazaurec², Guillermo Gallardo-Diez¹, and Rachid Deriche¹

¹ Athena EPI, Inria Sophia Antipolis-Méditerranée, Sophia Antipolis 06902, France

² ABS EPI, Inria Sophia Antipolis-Méditerranée, Sophia Antipolis 06902, France

Abstract. We present a graph-theoretical algorithm to extract the *connected* core structural connectivity network of a subject population. Extracting this core common network across subjects is a main problem in current neuroscience. Such network facilitates cognitive and clinical analyses by reducing the number of connections that need to be explored. Furthermore, insights into the human brain structure can be gained by comparing core networks of different populations. We show that our novel algorithm has theoretical and practical advantages. First, contrary to the current approach our algorithm guarantees that the extracted core subnetwork is connected agreeing with current evidence that the core structural network is tightly connected. Second, our algorithm shows enhanced performance when used as feature selection approach for connectivity analysis on populations.

1 Introduction

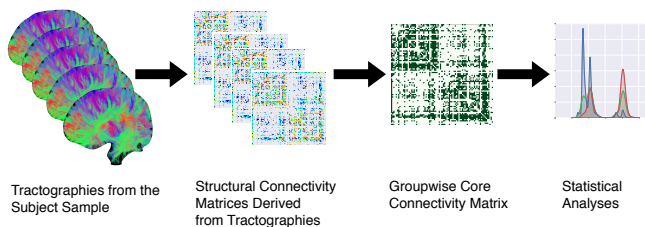


Fig. 1: Scheme of analyses involving the core structural connectivity matrix.

Isolating the common core structural connectivity network (SCN) of a population is an important problem in current neuroscience [3, 5]. This procedure facilitates cognitive and clinical studies based on Diffusion MRI [e.g. 1, 5] by increasing their statistical power through a reduction of the number of analyzed structural connections. We illustrate this process in Fig. 1. Furthermore,

recent evidence indicates a core common network exists in human and macaque brains and that it is tightly connected [2]. In this work we develop, for the first time, a group-wise core SCN extraction algorithm which guarantees a connected network output. Furthermore, we show the potential of such network to select gender-specific connections through an experiment on 300 human subjects.

The most used population-level core SCN extraction technique [5] is based on an effective statistical procedure to extract a population SCN: 1) computing, for each subject, a connectivity matrix using a standardised parcellation; and 2) extracting a binary graph by analysing each connection separately and rejecting hypothesis is not in the population. The resulting graph can be a set of disconnected subgraphs. This is problematic, recent studies have shown the core network to be tightly connected [2]. However, extracting connected group-wise core SCN is far from simple: an algorithm to find the largest core network of a population cannot find an approximated solution in polynomial time.

In this work, we propose a graph-theoretical algorithm to obtain the connected core SCN of a subject sample. Our approach guarantees a connected core SCN, agreeing with novel evidences on structural connectivity network topology [e.g. 2]. We start by proving that we can formulate the problem such that core network extraction is NP-Complete in general but in our case we find an exact polynomial time algorithm to perform the extraction. Finally, we show that our algorithm outperforms that of Gong et al. [5] as a tool for selecting regressors in SC. For this, we use 300 subjects from the HCP database and comparing the performance of the networks obtained with both algorithms to predict connectivity values from gender in a subsection of the core network.

2 Definitions, Problems, and Contributions

A first approach to core sub-network identification can be derived from the binary connectivity model. In this model the cortical and sub-cortical regions are common across subjects and what varies is whether these regions are connected or not [5]. Using this approach, a sample of human brain connectivity of a given population can be represented by $k \geq 1$ graphs $G_1 = (V, E_1), \dots, G_k = (V, E_k)$. In this formalism each graph G_i corresponds to a subject and, in accordance with Gong et al. [5], the vertices, V , stable across subjects are cortical and sub-cortical regions and the edges E_i are white matter bundles connecting those regions. Note that all graphs have the same ordered set of nodes. A first approach to compute, or approximate, the core sub-network of the population sample consists in finding the core SCN graph $G^* = (V^*, E^*)$ such that G^* and every G_i has some quantitative common properties, where $V^* \subseteq V$. In this article, we model the difference between the core SCN, G^* , and the subject ones, G_i , by a function f_λ . This function measures the difference between the sets of edges (and the sets of non-edges) of the core network and those of the subjects:

$$f_\lambda(G^*, G_i) = \lambda |\{e \in E, e \notin E(G_i[V^*])\}| + (1 - \lambda) |\{e \notin E, e \in E(G_i[V^*])\}|,$$

where $\lambda \in [0, 1]$ and $G_i[V^*]$ is the subgraph of G_i induced by the set of nodes $V^* \subseteq V$ and $|S|$ is the cardinality of a set S . In other words, f_λ represents the

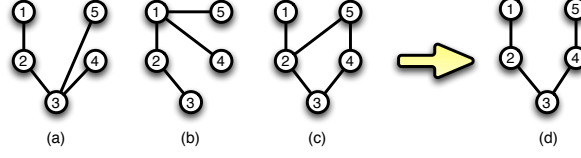


Fig. 2: Instance of the common sub-network problem. (a,b,c) Brain connectivity of different subjects, namely G_1 , G_2 and G_3 . (d) Extracted common sub-network G^* that is optimal for $n = 5$ with $\lambda = \frac{1}{2}$: the difference threshold is $\frac{7}{2}$.

difference between the set of edges of G and the set of edges of G_i modulated by the parameter λ . In the following, we will refer to $f_\lambda(G^*, G_i)$ as the *difference threshold* of a core sub-network G^* wrt G_i . Note that if $\lambda = 1$, we only consider edges excluded from the core network, $|\{e \in E, e \notin E(G_i[V^*])\}|$, and if $\lambda = 0$, we only consider edges included in the core network, $|\{e \notin E, e \in E(G_i[V^*])\}|$. In Definition 1, we formalize the problem of computing the core sub-network as a combinatorial optimization problem:

Definition 1 (Core sub-network problem). Let $G_1 = (V, E_1), \dots, G_k = (V, E_k)$ be $k \geq 1$ undirected graphs. Let λ be any real such that $\lambda \in [0, 1]$. Let $n \geq 0$ be any integer. Then, the core sub-network problem consists in computing a connected graph $G^* = (V^*, E^*)$ such that $|V^*| \geq n$ and such that the sum of the difference thresholds $\sum_{i=1}^k f_\lambda(G^*, G_i)$ is minimum.

Small example: Consider the instance depicted in Figure 2 with $\lambda = \frac{1}{2}$. Figure 2 (a), (b), and (c) represent G_1 , G_2 , G_3 , respectively. Figure 2 (d) is a solution $G^* = (V^*, E^*)$ when $n = 5$. Indeed, we have $f_\lambda(G^*, G_3) = \frac{1}{2}$ because the difference between G^* and G_3 is the edge connecting nodes 2 and 5 or the two element set $\{2, 5\}$; we have $f_\lambda(G^*, G_2) = 2$ because the difference between G^* and G_2 is the four edges $\{1, 5\}$, $\{1, 4\}$, $\{3, 4\}$, and $\{4, 5\}$; and we have $f_\lambda(G^*, G_1) = 1$ because the difference between G^* and G_1 is the two edges $\{3, 5\}$ and $\{4, 5\}$. We get $f_\lambda(G^*, G_1) + f_\lambda(G^*, G_2) + f_\lambda(G^*, G_3) = \frac{7}{2}$.

In the rest of this section we state our main contribution, an optimal polynomial time exact algorithm for the core sub-network problem if the number of nodes is sufficiently large (optimal means here that there is no exact algorithm with better complexity). Solving the problem in Def. 1, is hard: it can be proved that given an integer $n \geq 0$ and a real number $\delta \geq 0$, then the decision version of the SCN problem is NP-complete even if $k = 2$. However, focusing on the problem of minimizing f_λ we obtain a polynomial time algorithm for SCN extraction.

The main point of this work is to present an algorithm for the core graph extraction and assess its potential for clinical and cognitive studies. Even if the problem is very difficult to solve in general, we design our polynomial time core subnetwork extraction algorithm and show that it is optimal, when we focus on the problem of minimizing the difference threshold and when the number of nodes of the core sub-network is large.

Theorem 1. Consider $k \geq 1$ undirected graphs $G_1 = (V, E_1), \dots, G_k = (V, E_k)$ and consider any real number $\lambda \in [0, 1]$. Then, **Core-Sum-Alg** (Algorithm 1) is an $O(\max(k, \log(|V|)) \cdot |V|^2)$ -time complexity exact algorithm for the core sub-network problem when $n = |V|$.

Algorithm 1 Core-Sum-Alg: Exact polynomial time complexity algorithm for the core sub-network problem when $n = |V|$.

Require: SC graphs for each subject $G_1 = (V, E_1), \dots, G_k = (V, E_k)$, and $\lambda \in [0, 1]$

Start Computing a Core Graph that can have Multiple Connected Components:

- 1: Construct $\mathcal{G} = (V, V \times V)$, the completely connected graph \mathcal{G}
- 2: Compute $w_0(\cdot)$, $w_1(\cdot)$ across all subject graphs as in Eq. 1.
- 3: Compute the set of edges to be added to the core graph $E^1 = \{e | w_1(e) \leq w_0(e)\}$ and construct $\mathcal{G}_1 = (V, E^1)$.

Compute the Connected Components of \mathcal{G}_1 and Connect Them

- 4: Compute the set of maximal connected components $cc(\mathcal{G}_1) = (cc_1(\mathcal{G}_1) \dots cc_t(\mathcal{G}_1))$
 - 5: Construct $\mathcal{G}_{cc} = (\mathcal{V}_{cc}, \mathcal{V}_{cc} \times \mathcal{V}_{cc})$ with $\mathcal{V}_{cc} = \{u_1, \dots, u_t\}$
 - 6: Compute w_{cc} as in Eq. 4.
 - 7: Compute the set of edges E^0 that correspond to argument minimum of Eq. 4. In other words, for every $\{u_i, u_j\}$, select the edge e connecting the maximal connected components $cc_i(\mathcal{G}_1)$ and $cc_j(\mathcal{G}_1)$ such that $w_1(e) - w_0(e) = w_{cc}(\{u_i, u_j\})$.
 - 8: Compute a minimum spanning tree T_{cc} of \mathcal{G}_{cc}
 - 9: Compute the set of edges $E_*^0 \subseteq E^0$ that corresponds to the set of edges of the previous minimum spanning tree.
 - 10: Construct $G^* = (V^*, E^*)$ with $V^* := V$ and $E^* := E^1 \cup E_*^0$
 - 11: **return** G^* the connected Core Structural Connectivity Network
-

In the following, we aim at proving Theorem 1. Consider $k \geq 1$ undirected graphs $G_1 = (V, E_1), \dots, G_k = (V, E_k)$ and consider any real number λ such that $\lambda \in [0, 1]$. Let us define some notations and auxiliary graphs. Let $\mathcal{G} = (V, V \times V)$ be the completely connected cortical network graph with $V \times V$ the set all pairs from the elements from V . We define two edge-weighting functions w_0 and w_1 :

$$w_0(e) = (1 - \lambda)|\{i, e \in E_i, 1 \leq i \leq k\}|, \quad w_1(e) = \lambda|\{i, e \notin E_i, 1 \leq i \leq k\}|. \quad (1)$$

Intuitively, $w_0(e)$ represents the *cost* of not adding the edge e in the solution and $w_1(e)$ represents the *cost* of adding the edge e in the solution. From this, we define the graph induced by the set of edges to keep in the core subnetwork

$$\mathcal{G}_1 = (V, E^1), \quad E^1 = \{e \mid w_1(e) \leq w_0(e)\} \subseteq V \times V. \quad (2)$$

If \mathcal{G}_1 is a connected graph, then it is an optimal solution. Otherwise, we have to add edges in order to obtain a connected graph while minimizing the cost of such adding. To add such edges we define a graph representing the fully connected graph where each node represents a maximal connected component:

$$\mathcal{G}_{cc} = (\mathcal{V}_{cc}, \mathcal{E}_{cc}) \text{ with } \mathcal{V}_{cc} = \{u_1, \dots, u_t\} \text{ and } \mathcal{E}_{cc} = \mathcal{V}_{cc} \times \mathcal{V}_{cc}, \quad (3)$$

where $cc(\mathcal{G}_1) = (cc_1(\mathcal{G}_1), \dots, cc_t(\mathcal{G}_1))$ is the t maximal connected components of \mathcal{G}_1 . Then, to select which maximal connected components to include in our core subnetwork graph, we define a weight function w_{cc} :

$$w_{cc}(\{u_i, u_j\}) = \min_{v \in V(cc_i(\mathcal{G}_1)), v' \in V(cc_j(\mathcal{G}_1))} w_1(e) - w_0(e), \text{ where } 1 \leq i < j \leq t. \quad (4)$$

We formally prove in Lemma 1 that the problem of obtaining a minimum connected graph from \mathcal{G}_1 , that is solving the core sub-network problem when $n = |V|$, consists in computing a minimum spanning tree of \mathcal{G}_{cc} .

Lemma 1. *The core sub-network problem when $n = |V|$ is equivalent to compute a minimum spanning tree of $\mathcal{G}_{cc} = (V_{cc}, \mathcal{E}_{cc})$ with weight function w_{cc} .*

Proof. The core sub-network problem when $n = |V|$ consists in computing a graph $G^* = (V^*, E^*)$ such that $V^* = V$ and $\delta^* = \sum_{i=1}^k f_\lambda(G^*, G_i)$ is minimum. Consider the graph $\mathcal{G}_1 = (V, E^1)$ previously defined. Observe that: $\delta^* \geq \sum_{e \in V \times V} \min(w_0(e), w_1(e)) = \sum_{e \in E^1} w_1(e) + \sum_{e \in V \times V | e \notin E^1} w_0(e)$. Indeed, for every pair of nodes v, v' of V , either we set $\{v, v'\} \in E^*$ if $w_1(\{v, v'\}) \geq w_0(\{v, v'\})$ or we set $\{v, v'\} \notin E^*$. Thus, if \mathcal{G}_1 is a connected graph, then $G^* = \mathcal{G}_1$ is an optimal solution such that $\delta^* = \sum_{e \in V \times V} \min(w_0(e), w_1(e))$. Otherwise, we have to add edges in E^1 in order to get a connected graph (that is a spanning graph) and the “cost” of this addition has to be minimized.

Thus, suppose that the graph \mathcal{G}_1 contains at least two maximal connected components. Let $cc(\mathcal{G}_1) = (cc_1(\mathcal{G}_1), \dots, cc_t(\mathcal{G}_1))$ be the $t \geq 1$ maximal connected components of \mathcal{G}_1 . We have to connect these different components minimizing the increasing of the difference threshold. Let E^0 be the set of candidate edges constructed as follows. For every i, j , $1 \leq i < j \leq t$, let $\{v_i, v_j\}$ be an edge such that for every $v \in V(cc_i(\mathcal{G}_1))$ and for every $v' \in V(cc_j(\mathcal{G}_1))$, we have $w_1(\{v_i, v_j\}) - w_0(\{v_i, v_j\}) \leq w_1(\{v, v'\}) - w_0(\{v, v'\})$. In other words, $\{v_i, v_j\}$ is an edge that minimizes the marginal cost for connecting $cc_i(\mathcal{G}_1)$ and $cc_j(\mathcal{G}_1)$. We add $\{v_i, v_j\}$ in E^0 .

Thus, we have to add exactly $t - 1$ edges of E^0 in order to get a connected graph and we aim at minimizing the cost of this addition. More precisely, we get our optimal core network by finding, for every i, j , $1 \leq i < j \leq t$, an edge e such that $w_{cc}(\{u_i, u_j\}) = w_1(e) - w_0(e)$, that is an edge e of minimum marginal cost between the maximal connected component $cc_i(\mathcal{G}_1)$ and the maximal connected component $cc_j(\mathcal{G}_1)$. Let E_*^0 be such a subset of $t - 1$ edges. Thus, we get that $\delta^* = \sum_{e \in E^1} w_1(e) + \sum_{e \in V \times V | e \notin E^1, e \in E_*^0} w_1(e) + \sum_{e \in V \times V | e \notin E^1, e \notin E_*^0} w_0(e)$. Observe that $\sum_{e \in V \times V | e \notin E^1, e \in E_*^0} w_1(e)$ is exactly the cost of a minimum spanning tree of the graph \mathcal{G}_{cc} defined before. \square

We are now able to prove Theorem 1.

Proof.[of Theorem1] **Core-Sum-Alg** (Algorithm 1) follows the proof of Lemma 1 and so solves the core sub-network problem when $n = |V|$. The construction of \mathcal{G} (line 1) can be done in linear time in the number of edges, that is in $O(|V|^2)$ -time. The time complexity of line 2 is $O(k|V|^2)$. The construction of \mathcal{G}_1 (line 3)

can be done in $O(|V| + |E^1|)$ -time, $O(|V|^2)$ -time in the worst case. The computation of the maximal connected components of \mathcal{G}_1 (line 4) can be done in linear time in the size of \mathcal{G}_1 , that is in $O(|V| + |E^1|)$, $O(|V|^2)$ -time in the worst case. The construction of \mathcal{G}_{cc} (line 5) can be done in linear time in the size of \mathcal{G}_{cc} , that is, in the worst case, in $O(|V|^2)$ -time. The time complexity of line 6 and 7 is $O(|V|^2)$. There is an $O(m \log(n))$ -time complexity algorithm for the problem of computing a minimum spanning tree of a graph composed of n nodes and m edges (line 8). Thus, in our case, we get an $O(\log(|V|)|V|^2)$ -time algorithm. The time complexity of line 9 is $O(|V|)$. Finally, the construction of $G^* = (V^*, E^*)$ (line 10) can be done in constant time because $V^* = V$ and $E^* = E^1 \cup E_*^0$. \square

Having developed the core subnetwork extraction guaranteeing a connected core network (Algorithm 1). We proceed to assess its performance.

3 Experiments and Results



Fig. 3: Consistency analysis for extracted core graphs. We performed a Leave-N-Out procedure to quantify the consistency across methods at 4 different parameter levels. The results show similar graphs for both methods. However our method, in blue, has a smaller number of connections that are not present or absent across all experiments, i.e. unstable connections (marked in red).

To assess the performance of our method, we compared our novel approach with the currently used [5]: first, we compared the stability of the obtained binary graph across randomly chosen subpopulations; second, we compared connectivity prediction performance.

For this comparisons, we used an homogeneous set from the HCP500 dataset [6]: all subjects aged 21-40 with complete dMRI protocol, which resulted in 309 subjects (112 male). We obtained the weighted connectivity matrices between the cortical regions defined by the Desikan atlas [4] as done by Bassett

et al. [1]. To verify the untresholed graph construction, we computed the average degree, number connections over number of possible connections, on each subject. Bassett et al. [1] reported an average degree of 0.20 and we obtained 0.20 ± 0.01 (min: 0.17, max: 0.25) showing our preprocessing in agreement.

3.1 Consistency of the Extracted Graph

To quantify the consistency of the core graph extraction procedure we performed 500 Leave-N-Out experiments. At each experiment randomly sampled 100 subjects from the total and computed the core graphs with both techniques. We performed the extraction at 4 different levels of the parameter for each technique, choosing the parameters such that the density of the resulting graph connections is stable across methods. Also, we reported the number of *unstable connections*, selected as the connections that were not present or absent in all experiments. We show the results of this experiment in Fig. 3. In this figure we can observe that the resulting graphs are similar, while the number of unstable connections is larger for Gong et al. [5] by an order of magnitude.

3.2 Predicting Gender-Specific Connectivity

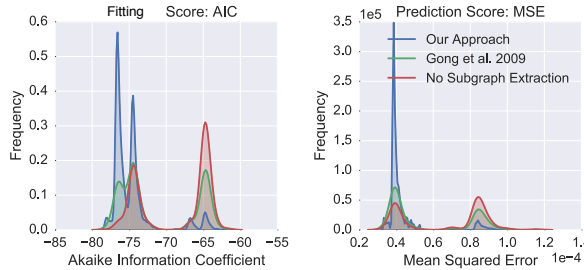


Fig. 4: Performance of core network as feature selection for a linear model for gender specific connectivity. We evaluate model fit (left) and prediction (right), Gong et al. [5] in green, and ours, in blue. We show the histograms of both values from our nested Leave- $\frac{1}{3}$ -Out experiment. In both measures, our approach has more frequent lower values, showing a better performance.

To assess model fit and prediction we implemented a nested Leave- $\frac{1}{3}$ -Out procedure. The outer loop performs model selection on $\frac{1}{3}$ of the subjects randomly selected. First, it computes the core graph of a population, with our approach and that of Gong et al. [5]. Then, it selects the features F that are more determinant of gender classification using the f-test feature selection procedure. The features are taken from the core graph adding the connectivity weights to each subject. The inner loop performs model fitting and prediction using the selected features F . First, we randomly take $\frac{1}{3}$ of the remaining subjects and fits a linear

model on F for predicting gender. Second, we predict the values of the features F from the gender column. The outer loop is performed 100 times and the inner loop 500 times per outer loop. This totals 50,000 experiments. Finally, for each experiment, we quantify the prediction performance of the linear model at each inner loop with the mean squared error (MSE) of the prediction and Akaike Information Criterion (AIC) for model fitting.

We show the experiment’s results in Fig. 4. In these results we can see that our approach, in blue, performed better than Gong et al. [5], in green as the number of cases with lower AIC and MSE is larger in our case.

4 Discussion and Conclusion

We present, for the first time, an algorithm to extract the core structural connectivity network of a subject population while guaranteeing connectedness. We start by formalizing the problem and showing that, although the problem is very hard (it is NP-complete), we produce a polynomial time exact algorithm to extract such network when its number of nodes is large. Finally, we show an example in which that our network constitutes a better feature selection step for statistical analyses of structural connectivity. For this, we performed a nested leave- $\frac{1}{3}$ -out experiment on 300 hundred subjects. The results show that performing feature selection with our technique outperforms the most commonly used approach.

Acknowledgments

This work has received funding from the European Research Council (ERC Advanced Grant agreement No 694665).

References

1. Bassett, D.S., Brown, J.A., Deshpande, V., Carlson, J.M., Grafton, S.T.: Conserved and variable architecture of human white matter connectivity. *NImg* (2011)
2. Bassett, D.S., Wymbs, N.F., Rombach, M.P., Porter, M.A., Mucha, P.J., Grafton, S.T.: Task-Based Core-Periphery Organization of Human Brain Dynamics. *PLoS Comput Biol* (2013)
3. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* (2009)
4. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M., Killiany, R.J.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NImg* (2006)
5. Gong, G., He, Y., Concha, L., Lebel, C., Gross, D.W., Evans, A.C., Beaulieu, C.: Mapping Anatomical Connectivity Patterns of Human Cerebral Cortex Using In Vivo Diffusion Tensor Imaging Tractography. *Cerebral Cortex* (2009)
6. Sotiropoulos, S.N., Jbabdi, S., Xu, J., Andersson, J.L., Moeller, S., Auerbach, E.J., Glasser, M.F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D.A., Yacoub, E., Lenglet, C., Van Essen, D.C., Ugurbil, K., Behrens, T.E.J.: Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NImg* (2013)